

## SAS, SPSS, and STATA Multilevel Modeling Syntax Guides

```

/*****
*****          SAS MULTILEVEL MODELING SYNTAX          *****
*****
*****/

/* PROC MIXED STATEMENT:
    DATA=:      File to use - default is last accessed
    NOCLPRINT:   Do not print class variable values
    NOITPRINT:   Do not print iteration history
    COVTEST:     Print SEs and p-values for significance test of variance estimates
    NAMELEN=:    # characters printed in fixed effects tables (default=20)
    METHOD=:      select REML or ML estimator - REML is default
    Other options...
    IC:          Print other information criteria and associated df
    MAXITER=:    # iterations (default=50)
    EMPIRICAL:   adjust SEs for non-normality of residuals (sandwich estimator) --
                not available with Satterthwaite or KR DDFM (use BW instead) */
PROC MIXED DATA=work.datafile NOCLPRINT NOITPRINT COVTEST NAMELEN=100 METHOD=REML;

/* CLASS statement makes SAS dummy code categorical variables (highest value is reference)
    Also include ID variables and REPEATED variables on CLASS statement */
CLASS IDvar catvar1 catvar2 cattime;

/* MODEL STATEMENT:
    Predict DV from time, catvar1, catvar2, contvar, 3 example interactions
    Don't need to construct interaction terms as variables a priori, use * between variables
    Options after the / ...
        NOINT:      Remove fixed intercept (is included by default)
        SOLUTION:   Print fixed effects solution (not included by default)
        DDFM=:      Change denominator degrees of freedom
                    Choose from Satterthwaite, KR, BW
    Other options...
        OUTP=:      Save predicted values from fixed+random effects to =dataset
        OUTPM=:     Save predicted values from fixed effects only to =dataset */
MODEL DV = time catvar1 catvar2 contvar catvar1*catvar2 time*catvar1 time*catvar2
/ SOLUTION DDFM=Satterthwaite;

/* RANDOM STATEMENT: Random Intercept must be listed if needed, also list any random slopes
    Options after the / ...
        G GCORR:    Print covariance and correlation matrices for random effects
        V VCORR:    Print covariance and correlation matrices for total outcome
                    V=1 VCORR=1 prints for 1st case (is default - can change if needed)
        TYPE=UN:    UN for unstructured to allow random effect covariances (VC=default)
        SUBJECT:    ID variable to identify nesting per level
    Other options...
        SOLUTION:   Print solution of random effects (the U's) - will take long time
        GROUP=:     =groupvar by which to get separate G matrices */
RANDOM INTERCEPT time / G GCORR V VCORR TYPE=UN SUBJECT=IDvar;

/* REPEATED STATEMENT: is always there even if not listed
    List variables repeated over (e.g., cattime) here and on CLASS statement
    Options after the / ...
        R RCORR:    Print covariance and correlation matrices for residuals
                    R=1 RCORR=1 prints for 1st case (is default - can change if needed)
        TYPE=:      VC for variance components (diagonal) by default
                    many, many other types available, such as AR(1), TOEP(n)
        SUBJECT:    ID variable to identify nesting per level (what is repeated over)
    Other options...
        GROUP=:     =groupvar by which to get separate R matrices
        LOCAL=:     EXP(predictor) for predictors of log of residual variance */
REPEATED cattime / R RCORR TYPE=VC SUBJECT=IDvar;

```

```

/* Execute PROC MIXED */ RUN;
/*****      OTHER OPTIONAL SAS PROC MIXED STATEMENTS      *****/

/* PARMs statement used to provide start values for variance components
   Must list as many ( ) as there are variance components in the model
   List in order of appearance in CovParms table
   Options after the / ...
       HOLD=:      means fix those components (here, #3 is fixed to 1)
       NOBOUND:    allows variances to go negative (useful for troubleshooting) */
PARMS (5) (3) (2) (1) / HOLD=3 NOBOUND;

/*      LSMEANS Generates means & tests for specified variables
       SLICE Tests effect of catvar1 at each level of catvar2
       if the time variable is at 1, uses Tukey Adjustment
       DIFF=ALL requests all possible pairwise comparisons */
LSMEANS catvar1*catvar2 / SLICE=catvar2 AT time=1 ADJUST=TUKEY DIFF=ALL;

/*      ESTIMATE is used for specific hypothesis tests
       "LABEL" is first, followed by effect being estimated.
       Below we ask for group differences in main effect of catvar1
       and group differences in the linear slope of time */
ESTIMATE "L vs. H Catvar1 for Main Effect"    catvar1 -1 0 1 ;
ESTIMATE "M vs. H Catvar1 for Main Effect"    catvar1 0 -1 1 ;
ESTIMATE "L vs. M Catvar1 for Main Effect"    catvar1 -1 1 0 ;
ESTIMATE "L vs. H Catvar1 for Linear Slope"    time*catvar1 -1 0 1 ;
ESTIMATE "M vs. H Catvar1 for Linear Slope"    time*catvar1 0 -1 1 ;
ESTIMATE "L vs. M Catvar1 for Linear Slope"    time*catvar1 -1 1 0 ;

/*      Below we ask for simple effects at the interacting variable=3 */
ESTIMATE "Simple Effect of X if Z=3"    xvar 1    xvar*zvar 3;
ESTIMATE "Simple Effect of Z if X=3"    zvar 1    xvar*zvar 3;

/*      ODS OUTPUT is used to save output tables to SAS datasets
       SolutionF=:      Save fixed effects to =dataset
       SolutionR=:      Save random effects to =dataset
       CovParms=:       Save covariance parameters to =dataset
       FitStatistics=:   Save fit statistics to =dataset
       Estimates=:       Save requested estimates to =dataset */
ODS OUTPUT    SolutionF=work.FixedEffects
               CovParms=work.CovarianceParameters
               FitStatistics=work.FitStats;

*****
*****      SPSS MULTILEVEL MODELING SYNTAX      *****
*****

* MIXED STATEMENT: DV is listed first
  BY:    list categorical predictors (main effects only)
  WITH:  list continuous predictors (main effects only).

* /METHOD=REML: used to select estimator (REML is default, ML is also available).

* /PRINT=: used to request specific output
  SOLUTION:      Print fixed effects solution (not default)
  TESTCOV:       Print SEs and p-values for significance tests for variances (not default)
  G:             Covariance matrix of random effects (no G correlation matrix available)
  R:             Covariance matrix of residuals (no R correlation matrix available)
  CPS:          Case processing summary: factor values, repeated measures variables,
               repeated measure subjects, random effects subjects & frequencies
  DESCRIPTIVES:  Sample sizes, means, SD of DV & covariates for each combination of factors
  HISTORY(1):    Iteration History (1=print every iteration).

```

```

* /FIXED=: used to specify fixed effects (intercept included by default)
    Don't have to define interaction terms ahead of time, can do so via * between variables
    Options after the | :
        SSTYPE(3): Sums of Squares Type (3=default, also 1 available)
        NOTINT: To remove fixed intercept.

* /RANDOM=: used to specify random effects (intercept NOT included by default)
    Options after the | :
        COVTYPE(UN)=: UN for unstructure to allow random effects covariances (not default)
        SUBJECT(IDvar): ID variable to identify nesting per level.

* /REPEATED=: is always there even if not listed
    List variables repeated over (e.g., cattime) here and on BY statement
    Options after the | :
        COVTYPE(DIAG)=: DIAG for diagonal (default)
        SUBJECT(IDvar): ID variable to identify nesting per level (what repeated over).

***** OTHER OPTIONAL MIXED STATEMENTS *****

* /CRITERIA: used to change estimation options (leave the convergence ones alone)
    MXITER(100): Change number of iterations (100=default).

* /SAVE=: used to save predicted values to dataset
    FIXPRED: Save estimates for predicted values from fixed effects only
    SEFIXP: Save SE for predicted values from fixed effects only
    DFFIXP: Save Satterthwaite DDFM for predicted values from fixed effects only
    PRED: Save estimates for predicted values from fixed+random effects
    SEPRED: Save SE for predicted values from fixed+random effects
    DFPRED: Save Satterthwaite DDFM for predicted values from fixed+random effects
    RESID: Save residuals from fixed+random effects.

* /EMMEANS: used to request means for categorical predictors
    TABLES: list variables to get means for (unique combination)
    WITH: values of continuous predictors to be evaluated at
    COMPARE(var): variable to be compared per level of interacting variable
    ADJ(LSD): pairwise comparison (LSD=no adjustment, also Bonferroni, Sideak).

* /TEST is used for specific hypothesis tests
    "LABEL" is first, followed by effect being estimated
    Below we ask for group differences in main effect of catvar1
    and group differences in the linear slope of time
    Next we ask for simple effects at the interacting variable=3.

MIXED DV BY IDvar catvar1 catvar2 WITH time contvar
/METHOD = REML
/PRINT = SOLUTION TESTCOV G R
/FIXED = catvar1 catvar2 contvar catvar1*catvar2 time*catvar1 time*catvar2
/RANDOM = intercept time | COVTYPE(UN) SUBJECT(IDvar)
/REPEATED = cattime | COVTYPE(DIAG) SUBJECT(IDvar)
/* Other optional commands would follow */
/EMMEANS TABLES(catvar1*catvar2) WITH(time=0) COMPARE(catvar1) ADJ(LSD)
/* Examples of TEST commands */
/TEST = "L vs. H Catvar1 for Main Effect" catvar1 -1 0 1
/TEST = "M vs. H Catvar1 for Main Effect" catvar1 0 -1 1
/TEST = "L vs. M Catvar1 for Main Effect" catvar1 -1 1 0
/TEST = "L vs. H Catvar1 for Linear Slope" time*catvar1 -1 0 1
/TEST = "M vs. H Catvar1 for Linear Slope" time*catvar1 0 -1 1
/TEST = "L vs. M Catvar1 for Linear Slope" time*catvar1 -1 1 0
/TEST = "Simple Effect of X if Z=3" xvar 1 xvar*zvar 3
/TEST = "Simple Effect of Z if X=3" zvar 1 xvar*zvar 3.

```

```

*****
*****          STATA MULTILEVEL MODELING SYNTAX          *****
*****

* GENERIC EXAMPLE SYNTAX FOR XT MIXED in STATA 12 or MIXED in STATA 13:
  xtmixed DV fixed effects, FE options || Level2ID: random effects, RE options ///
    variance reml/mle covariance(Gmatrixtype) residuals(Rmatrixtype, t(Level1ID)),
  estat ic n(#persons),

* Fixed effects (FE) options:
  * noconstant to remove fixed intercept (included by default)
  * i. indicates categorical predictors (reference is first by default)
  * c. indicates continuous predictors (default if not specified)
  * can fit interactions on the fly
    * c.age#c.age creates quadratic age slope
    * i.group#c.age creates group by age interaction
    * i.program##i.day creates program by day categorical interaction

* Random effects (RE) options:
  * noconstant to remove random intercept (included by default)
  * covariance(Unstructured) is for G matrix unstructured
  * estat recovariance --> display G matrix
    * options: , level(levelvar) correlation
    * levelvar says at what level, correlation prints GCORR
  * Can add another level as || Level3ID: random effects, RE options
  * Can add group predictors of random effects heterogeneity
    * gen: boyXage = boy*age
    * ID: boy boyXage, no constant --> separate int and age slope per gender
  * Add R. to indicate categorical variables as random effects
  * Add _all: instead of ID: to indicate no nesting
  * Can do crossed models
    * || _all: R.id || _all: R.week --> persons by weeks as crossed
    * More efficient and equivalent: || _all: R.week || id:

* Residual options:
  * R matrix via residual(form,options)
    * independent-->VC, exchangeable-->CS, unstructured, toeplitz #-->TOEPn
    * AR #, exponential --> AR for unbalanced time
  * option by(varname) allows heterogeneous residual variance
  * option t(varname) is level-1 ID variable (i.e., for time)

* General options:
  * mle for ML, reml for REML is default, variance asks for variances rather than SD
  * noretale for no random effects solution, nogroup for no table summarizing groups
  * noheader suppresses output header, estat ic prints AIC and BIC (#parms = #total parms)

*** Other options

* estimates --> Can store results and do LR test comparisons
* lrtest --> LR test for models listed that have been saved
  estimates store bigmodel, estimates store smallmodel
  lrtest bigmodel smallmodel
* predict --> predicted estimates, linear predictor and SE from fixed effects
  predict xb
* lincom --> point estimates and SEs for linear combinations (like ESTIMATE)
  lincom 1*xvar + 3*xvar*zvar          // Simple effect of X if Z=3
  lincom 1*zvar + 3*xvar*zvar          // Simple effect of Z if X=3
* margins --> marginal means (LSMEANS)
* Estimating group means at first and last occasions
  margins ib(last).catvar1, at(c.time=(0) c.timesq=(0 ))
  margins ib(last).catvar1, at(c.time=(5) c.timesq=(25))

* test --> Wald test of simple and composite linear hypotheses
* Example contrasts between groups on intercept and linear time slopes
  test 1.catvar1=3.catvar1              // Low vs. High: Intercept
  test 2.catvar1=3.catvar1              // Med vs. High: Intercept
  test 1.catvar1=2.catvar1              // Low vs. Med: Intercept
  test 1.catvar1#time=3.catvar1#time    // Low vs. High: Linear Slope
  test 2.catvar1#time=3.catvar1#time    // Med vs. High: Linear Slope
  test 1.catvar1#time=2.catvar1#time    // Low vs. Med: Linear Slope

```